

Multi-Classify Hybrid Multilayered Perceptron (HMLP) Network for Pattern Recognition Applications

Fakroul Ridzuan Bin Hashim, John J. Soraghan & Lykourgos Petropoulakis
Centre for Excellent in Signal and Image Processing,
University of Strathclyde, Department of Electronic & Electrical Engineering,
204 George St, Glasgow G1 1XW, United Kingdom.
fakroul.bin-hashim@strath.ac.uk

Abstract-This paper introduces a Hybrid Multilayered Perceptron (HMLP) based classifier known as the Multi-Classify HMLP network (MCHMLP). This network is shown to be able to enhance the performance accuracy when compared to the conventional HMLP network. The Multi-Classify HMLP network architecture is trained using a Modified Recursive Prediction Error (MRPE). This study uses three benchmark datasets in order to measure the capability of the network. The results show that the proposed Multi-Classify HMLP network provides a significant improvement over the conventional HMLP network for pattern recognition applications.

I. INTRODUCTION

Artificial neural networks (ANNs) can be used to solve problems without the need to create a model of a real biological system. ANNs pattern recognition and classification units are widely used in medicine, banking analysis, stock market prediction, telecommunication industry and many other technology and science areas. Mat-Isa *et al.*, [1] have shown that neural networks are able to identify the risk level in patients with pre-cancerous cervix. Neural networks are used to predict the stock market to ensure that profits can be enhanced or losses can be minimized [2]. Liu *et al.* use neural networks as the prediction agent in measuring the earthquake tremor levels that can occur based on the identified features. In [3] ANNs have been used in aiding earthquake predictions along with the related tremors analysis. In another study, Mat-Isa *et al.* showed that ANNs can provide good results in identifying the shape of aggregates. The aggregates are identified and separated into groups before being used in construction and those aggregates with good shape are selected to ensure the strength of the concrete at the highest level [4].

Multilayered Perceptron (MLP) neural network is a conventional example of an ANN network. There are several modified versions of the MLP networks. The Hybrid MLP (HMLP) network is one such version, which has added connections to feed information from the input layer directly to the output layer. Mashor [5] has shown an increased accuracy of results provided by the HMLP network compared to the conventional MLP networks. The ability to make predictions of the neural network is dependent on network training. There are many types of training algorithms for training neural networks to increase prediction. Popular training algorithms include Back Propagation (BP), Lavenberg Marquardt (LM), Bayesian Regularization (BR), Recursive Least Squares (RLS), and

Recursive Prediction Error (RPE). Mashor in [6] has shown that the accuracy could be improved if some modifications were made to the RPE training algorithms. The Modified Recursive Prediction Error (MRPE) is the modified version of the RPE training algorithm which can increase the accuracy of predictions of the conventional RPE algorithm when training the HMLP network. He also stated that the MRPE training algorithm is better suited for use with the HMLP network as it resulted in more accurate predictions than other approaches.

In this study, modifications to the HMLP network will be made to improve the performance of the resulting network. The presented method proposes that the classification process is done repetitively in multi-stages. The output outcomes from a HMLP network in the first stage will be presented as input to another HMLP network in the second stage. Courtney *et al.* [7] in their research used a multi-stage neural network to measure the colour constancy and the colour induction using 'NEXUS'. Nexus [8] is a large scale interactive neural simulator designed to simulate a huge number of networks with approximately 1.65 million connections. Lai *et al.* [9] used multi-stage neural networks to predict the foreign exchange rate (FOREX). Multiple training algorithms were used to train the network before the pruning phase took place in order to produce the selected base models. A multi-stage neural network was implemented to recognise the pattern of a mammogram screen [10]. Back propagation with a Kalman filter was applied as the training algorithm to perform the classification.

In the work presented in this paper three datasets, taken from the University of California – Irvine (UCI) repository [11] are used in order to investigate whether the proposed Multi-Classify HMLP network can provide more accurate results than conventional networks. These were: the Breast Cancer Wisconsin (BCW) dataset, the Liver Disorder (LD) dataset the Pima Indians Diabetes (PID) dataset. The remainder of the paper is organized as follows. In section II the proposed method is shown in detail followed by the results and discussion in section III. Conclusions are presented in section IV.

II. MULTI-CLASSIFY HYBRID MULTILAYERED PERCEPTRON NETWORK

It is well known [5] that modelling a linear framework using a standard nonlinear MLP network is not a good decision. For this reason a hybrid version of the MLP network was introduced known as the HMLP network. This involved adding linear connections linking the input

layer directly to the output layer without going through the hidden layer, an approach which proved to significantly improve performance. Thus, the HMLP network was shown to be sufficiently competent for designing optimal structures both for linear and nonlinear systems. By directly linking the network input to the output node through a number of weighted connections, HMLP allows the formation of a linear system. Additional connections can be made as represented by the dashed line shown in Figure 2.

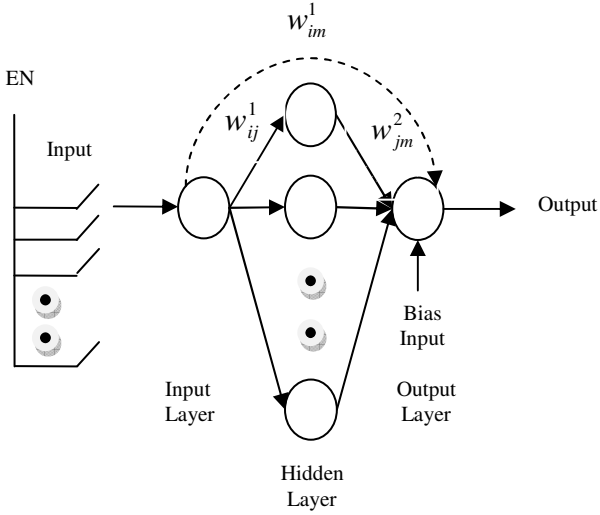


Fig. 2: A schematic diagram of HMLP network with one hidden layer.

For the output node, the output of the network is given by:

$$\hat{y}(t) = \sum_{j=1}^{n_h} w_{jk}^2 F \left(\sum_{i=1}^{n_i} w_{ij}^1 x_i^0(t) + b_j^0 \right) + \sum_{i=1}^{n_i} w_{ik}^1 x_i^0(t) \quad (1)$$

for $1 \leq j \leq n_h$ and $1 \leq k \leq m$

where w_{ik}^1 denotes the weights of the additional linear connection between the input and output layer, n_h is the hidden node while m is the number of outputs of the network.

In this study we propose a modified version of the HMLP network that is formed by repeating the classification process in a second stage. The proposed method is affected by adding another HMLP network to the output layer of the conventional HMLP network.

Figure 3 shows the block diagram of the architecture of the proposed classification method using two HMLP networks. These networks are in cascade so that the output of the first network is the input of the second network. The first HMLP network is fed with training input data. Each data consists of several parameters representing the data features. These features correspond to the various attributes used for the classification in each case. Detailed explanation of these attributes is given in

section III of the paper. Here the network was trained in order to find the best performance of the classification process.

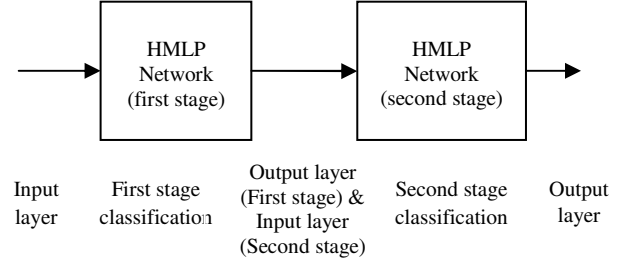


Fig. 3: Block diagram of the proposed network architecture.

It is considered that the best possible performance has been attained when the error of the network has reached a small value and remains constant even though the number of nodes increases. At that moment it is considered that the network has obtained the highest level of convergence. Several statistical techniques are available to represent the error of system such as Mean Square Error (MSE), Sum Square Error (SSE), and Standard Deviation (SD). In this case the Standard Deviation is used to represent the error of the network.

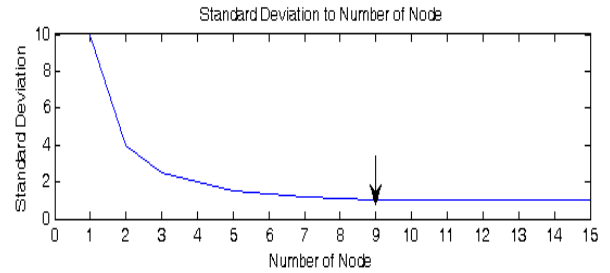


Fig. 4: A graph of Standard Deviation variation with increasing Number of Network Nodes

Figure 4 shows that the smallest SD occurred after using 9 nodes (arrow). After this point the value of SD remains unchanged although the number of nodes increases. With the highest convergence reached, a set of test data is provided to test the HMLP network. A good trained network is indicated by the ability to represent the whole dataset at the training phase. Once the best performance has been achieved for the first HMLP its outcome is then used as input to the second stage HMLP network. The same training and testing procedures applied to the first HMLP network are also used to train and test the second stage HMLP network. In this case only one feature is represented for each data input since the first stage HMLP network provides only one feature at the output layer as outcome.

Several aspects of the network training must be fulfilled in order to obtain results with the highest possible accuracy. The type of network structure must be matched

by an appropriate training algorithm and an adequate size of data must exist to train the network, thus avoiding poor results. Neural networks cannot tolerate small size datasets because they are trained by the history of the dataset and try to adjust their weights to best fit the supplied data. The objective of applying the multi-stage classification is to redo the classification that happens at the first stage network and provide the second stage network an opportunity to improve the overall result. Each neural network has its own convergence limitations based on the structure of the network, type of training algorithm used and the size of the data available for training. If the accuracy of the network keeps giving the same results even when the number of iterations is increased, this means the network is near or at its highest convergence point. So, by using the output of the first stage as input for the second network we could improve on the convergence limitations of first HMLP network. The second stage network will re-classify the wrong classification held at the first stage network. The function of the second HMLP network is to automatically readjust the classification thresholds thus improving the performance obtained compared to a single HMLP.

III. BENCHMARK DATASETS

The performance of the proposed MCHMLP network using MRPE as the training algorithm is evaluated by comparing it with a multi-stage MLP network, and conventional HMLP, MLP, HRBF and RBF networks. The 3 benchmark datasets from the UCI machine learning repository [11], i.e. the Breast Cancer Wisconsin, the Liver Disorder, and the Pima Indians Diabetes, have been used to examine the performance of the MCHMLP network. The results are then compared to the results from the aforementioned networks.

The Breast Cancer Wisconsin dataset contains a set of subjects with breast cancer and characterized by 9 attributes. The subjects have been diagnosed either with benign tumors or malignant tumors. The dataset has 699 entries with 458 benign tumors and 241 malignant tumors. The 9 attributes are: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nucleoli, bland chromatin, normal nucleoli and mitoses. The dataset is divided into two groups with 400 comprising the training dataset while 299 make up the testing dataset for the neural networks.

The Liver Disorder dataset classifies the existence of liver disorder in subjects with active alcoholic activity. Blood samples are tested for each subject and, according to the results the subject is placed in group 1 or in group 2. The dataset has 345 entries and is divided into two groups with 145 entries in group 1 and 200 entries in group 2. The 6 attributes are used to classify the subjects into these groups are: mean corpuscular volume (mcv), alkphos alkaline phosphotase, alamine aminotransferase (sgpt), aspartate aminotransferase (sgot), gamma-glutamyl transpeptidase (gammagt) and drinks number of half-pint equivalents of alcoholic beverages drunk per day. For the purpose of training and testing the neural networks, this

dataset is divided into two subsets with 200 entries for the training dataset and 145 for the testing dataset.

The Pima Indians Diabetes dataset is taken from female subjects with at least 21 years old of Pima Indians heritage. The subjects have been diagnosed either as healthy or as diabetics. The dataset has 8 attributes: number of pregnancies, plasma glucose concentration, diastolic blood pressure, triceps skin fold thickness, 2-h serum insulin, body mass index, diabetes pedigree, and last but not least, the age of the subject. The database has 768 entries with 268 having tested positive for diabetes. For the purpose of training and testing the neural networks, this dataset is divided into a training set with 400 entries and a testing set with 368 entries.

IV. EXPERIMENTAL RESULTS

Several types of networks have been used and compared in this study using the 3 datasets mentioned earlier. The MCHMLP along with the multi-stage MLP, the HMLP, the Hybrid Radial Basis Function (HRBF), the MLP, and the Radial Basis Function (RBF) networks were used to classify all three datasets. Training was carried out using the following techniques: the proposed MCHMLP, the multi-stage MLP, the HMLP and the MLP networks were trained using the MRPE training algorithm to obtain the data classification. The HRBF and the RBF networks were trained using the Given's Least Square (GLS) for clustering the dataset.

Table 1: Neural Networks performance using the Breast Cancer Wisconsin dataset for the training and testing phases.

Network	Accuracy, (%)		Number of Nodes	Standard Deviation, (SD)
	Training Phase	Testing Phase		
MCHMLP	97.25	97.03	15(fs) 6(ss)	2.32(fs) 0.93(ss)
Multi-stage MLP	94.50	92.30	14(fs) 2(ss)	2.48(fs) 0.95(ss)
HMLP	95.00	94.66	15	2.32
MLP	81.75	80.63	14	2.48
HRBF	82.00	76.58	1	5.20
RBF	75.25	73.58	1	5.92

*first stage (fs), second stage (ss).

Table 2: Neural Networks performance using the Liver Disorder dataset for the training and testing phases.

Network	Accuracy, (%)		Number of Nodes	Standard Deviation, (SD)
	Training Phase	Testing Phase		
MCHMLP	77.00	75.86	9(fs) 2(ss)	4.69(fs) 3.76(ss)
Multi-stage MLP	71.00	70.34	5(fs) 3(ss)	6.30(fs) 3.96(ss)
HMLP	74.50	73.17	9	4.69
MLP	69.00	67.43	5	6.30
HRBF	68.00	60.00	2	6.50
RBF	44.50	40.00	2	6.70

*first stage (fs), second stage (ss).

Table 3: Neural Networks performance using the Pima Indians Diabetes dataset for the training and testing phases.

Network	Accuracy, (%)		Number of Nodes	Standard Deviation, (SD)
	Training Phase	Testing Phase		
MCHMLP	77.50	77.05	12(fs) 4(ss)	3.68(fs) 3.06(ss)
Multi-stage MLP	65.25	64.90	14(fs) 6(ss)	3.78(fs) 3.74(ss)
HMLP	73.50	72.98	12	3.68
MLP	64.75	62.57	14	3.78
HRBF	58.00	51.52	2	5.92
RBF	56.00	51.52	2	4.85

*first stage (fs), second stage (ss).

Tables 1, 2 and 3 show the performance results of the neural networks using the three datasets. It is clear from these tables that the MCHMLP outperforms all the other networks, obtaining the highest overall accuracy for every dataset used. The performances of the multi-stage MLP, the HMLP and the MLP networks are quite good, however they are significantly lower than the performance of the MCHMLP. The results clearly show that when additional networks are added to the conventional HMLP and MLP structures the accuracy of the results increases considerably for a relatively small increase in the number of network nodes due to the introduced second stage. The RBF family of networks (RBF and HRBF) are able to classify the three datasets but no better than the MLP family.

The standard deviation in this study is used as the stopping criterion of network iteration to measure the convergence level. The MCHMLP shows the highest convergence and accuracy performance for the BCW dataset by using 15 and 6 nodes at the first and second stage networks with 2.32 and 0.93 of SD respectively. For the Liver Disorder dataset, the f network needed 9 nodes for the first and 2 nodes for the second stages. In the case of the PID dataset 12 and 4 nodes were required in each stage respectively. From the tables of the results it is obvious that the MLP family needs a larger number of nodes to reach the highest convergence when compared to the RBF family. Billings et al. in their study state the RBF family is able to perform faster than MLP networks, where the weight can be pre-fixed since they involve only linear optimization [12]. Although the RBF networks may be able to produce faster training sessions, are unable to produce as accurate results. This is well illustrated in our test cases. These results are also supported by Mat-Isa et al. who, in their research, found that the performance of the RBF family of networks cannot be better than that of the MLP family [1].

The results in Tables 2 and 3 also indicate that the MCHMLP offers superior performance however the difference in performance is not as clear as in Table 1. In the case of Liver Disorder (Table 2) and Pima Indian Diabetes (Table 2) problem occurs when there is a large overlap between two groups in the datasets. The datasets are divided into groups A and B but, parts of the data belong to either group A or group B. So, datasets with a high overlap lead to a difficult classification process

compared to datasets with a small overlap. The MCHMLP network is shown to be able to obtain the best classification results under all tested conditions in this study compared to the other networks. Nonetheless, datasets with a high overlap reduce the accuracy of the classification. However, Tables 1, 2 and 3 shows the MCHMLP network is capable to generalize the datasets in training phase and performs good results in testing phase.

V. CONCLUSION

In this study, the multi-classify HMLP network trained with MRPE as the training algorithm has been proposed for pattern recognition applications. The performance of the proposed network was measured by using three benchmark datasets. The results show that the multi-classify HMLP network is able to classify the Breast Cancer Wisconsin, Liver Disorder and Pima Indian Diabetes with significant accuracy (and in some cases with very high accuracy) and, in all cases, the performance was better than any of the other networks used in this study. Moreover, in all three cases examined, the introduction of the second stage network resulted in improved standard deviation results at the second stage, a direct indication of improved convergence of the network. The additional overhead of extra nodes and increased training time was not significant.

REFERENCES

- [1] N. A. Mat-Isa, M. Y. Mashor and N. H. Othman, "An automatic cervical pre-cancerous diagnostic system," *Journal of Artificial Intelligent in Medical*, vol. 42(1), pp. 1-11, 2008.
- [2] M. P. Naeini, H. Taremiyan, and H. B. Hashemi, "Stock market value prediction using neural networks," *International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, pp. 132 – 136, 2010.
- [3] Y. Liu, H. Liu, B. Zhang and G. Wu, "Extraction of if-then rules from trained neural network and its application to earthquake prediction," *Proceedings of the Third IEEE International Conference on Cognitive Informatics*, pp. 109-115, 2004.
- [4] N. A. Mat-Isa, M. S. Al-Batah, K. Z. Zamli, K. A. Azizi, A. Joret and N. R. Mat-Noor, "Suitable features selection for the HMLP and MLP networks to identify the shape aggregate," *Journal of Construction and Building Materials*, vol. 22(3), pp. 402-410, 2008.
- [5] M. Y. Mashor, "Hybrid multilayered perceptron networks," *International Journal of System Sciences*, vol. 31(6), pp. 771-785, 2000.
- [6] M. Y. Mashor, "Modified recursive prediction error algorithm for training layered neural network," *International Journal of The Computer, The Internet and Management*, vol. 11(2), pp. 29-36, 2003.
- [7] S. M. Courtney, L. H. Finkel and G. Buchsbaum, "A multistage neural network for color constancy and color induction," *IEEE Transactions on Neural Networks*, vol. 6(4), pp. 972-985, 1995.
- [8] P. Sajda and L. H. Finkel, "NEXUS: A simulation environment for large-scale neural systems," *SIMULATION*, 1992. ISSN: 00375497/92.
- [9] K. K. Lai, L. Yu, W. Huang and S. Wang, "Multistage neural network metalearning with application to foreign exchange rates forecasting," *Learning Notes in Artificial Intelligent*, vol.4293, pp. 338 – 347, 2006.

- [10] B. Zheng, W. Qian and L. P. Clarke, "Multistage neural network for pattern recognition in mammogram screening," IEEE International Conference on Neural Networks, IEEE World Congress on Computational Intelligence, vol. 6, pp. 337-344, 1994.
- [11] D. J. Newman, S. Hettich, C. L. Blake, C. J. Merz and D. W. Aha, "UCI Repository of machine learning database," Department of Information and Computer Science, University of California, Irvine, 1998, <http://archive.ics.uci.edu/ml/databases.html>, last accessed on March 10th 2012.
- [12] S. A. Billings, H. L. Wei and A. Balikhin, "Generalized multiscale radial basis function networks," *Journal of neural Network*, vol. 20, pp. 1081-1094, 2007.